보건학 석사학위 논문

Breast Cancer Risk Prediction via Mammographic Density Quantification in Digital Mammograms Using Deep Learning

유방 촬영술 영상 자료의 딥러닝 적용을 통한 유방암 위험도 평가 : 유방 치밀도 자동 평가 방법 기반

2019 년 5 월

서울대학교 보건대학원
보건학과 유전체 & 건강빅데이터 전공

안주영

Breast Cancer Risk Prediction via Mammographic Density Quantification in Digital Mammograms Using Deep Learning

유방 촬영술 영상 자료의 딥러닝 적용을 통한 유방암 위험도 평가 : 유방 치밀도 자동 평가 방법 기반

지도교수 성주헌

이 논문을 보건학 석사 학위논문으로 제출함

2019년 5월

서울대학교 보건대학원

보건학과 유전체 & 건강빅데이터 전공

안주영

안주영의 석사 학위논문을 인준함

2019년 6월

위 원 장 김호 (인)

부 위 원 장 조성일 (인)

위 원 성주헌 (인)

Abstract

Introduction : Mammographic density adjusted for age and body mass index (BMI) is the most predictive marker of breast cancer after familial causes and genetic markers. The aim of this study was to develop deep learning (DL) algorithm to assess mammographic density.

Methods : Total 2464 participants (834 cases and 1630 controls) were collected from Asan Medical Center and Samsung Medical Center, Korea. Cranio-caudal view mammographic images were obtained using full-field digital mammography system. Mammographic densities were measured using CUMULUS software. The resulting DL algorithm was tested on a held-out test set of 493 women. Agreement on DL and expert was assessed with correlation coefficient and weighted κ statistics. Risk associations of DL measures were evaluated with area under curve (AUC) and odds per adjusted standard deviation (OPERA).

Results : The DL model showed very good agreement with expert for both percent density and dense area ($r = 0.94 - 0.96$ and $\kappa = 0.89 - 0.91$). Risk associations of DL measures were comparable to manual measures of expert. DL measures adjusted for age and BMI showed strong risk associations with breast cancer (OPERA = 1.51 - 1.63 and AUC = 0.61 - 0.64).

Conclusions : DL model can be used to measure mammographic density which is a strong risk factor of breast cancer. This study showed the potential of DL algorithm as a mammogram-based risk prediction model in breast cancer screening test.

Keywords: Mammographic density, Breast cancer, Risk prediction, Deep learning, Digital mammogram

Student Number: 2017-27525

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Mammographic features, which represent the radiological appearance of breast are associated with breast cancer risk [1-5]. Mammographic density, one of the major mammographic features, has conventionally been defined as bright areas on a mammogram, which reflects amount of fibroglandular tissue in a breast [2, 3]. Once adjusted for age and body mass index (BMI), mammographic density is the most predictive marker of breast cancer after familial causes and genetic markers [4, 5].

Mammographic density has not been integral part of clinical decision making for breast cancer preventive intervention [6]. The key challenge of integrating mammographic density in clinical practice was that it requires a lot of time and cost to measure. Measuring mammographic density was also dependent on a reader's subject interpretation on a mammogram, which may cause measurement error due to intra- and inter-radiologist variability. To handle these issues, several automatic programs were presented, however concerns about validity have been consistently raised, because technical parameters such as compression and radiation dose were expected to affect measurement [7].

Deep learning (DL) have been successfully applied to the medical image interpretation. There have been previous studies investigating DL for mammographic density assessment [9-12]. Previous studies mainly focused on qualitative assessment of mammographic density based on Breast Imaging Reporting and Data System (BI-RADS) category [11, 12].

Several studies proposed DL method to quantify mammographic density [9, 10]. Kallenberg et al. presented unsupervised deep learning approach called convolutional sparse auto-encoder to extract features from mammogram and tried to measure mammographic density with following supervised neural network [9]. Lee et al. adopted fully convolutional network to segment fibroglandular tissue and quantify mammographic density [10]. However, little work has been done on evaluating automated measures with respect to breast cancer risk prediction in consideration of clinical information.

The aim of this study was to develop DL model which can be used to assess mammographic density and breast cancer risk in clinical practice. DL model was developed with extensive digital mammograms from two institutions, for which mammographic density was quantitatively measured by an expert. The resulting DL model was tested whether it can discriminate breast cancer cases from controls as much as an expert.

Chapter 2

Materials and Methods

2.1 Data collection

Total 2464 participants (834 cases and 1630 controls) were collected for this study. First part of participants were from Asan Medical Center, Seoul, Korea (728 cases and 1375 controls). All cases from this institution were women who were diagnosed with breast cancer (ductal carcinoma in situ or invasive breast carcinoma) and had undergone breast cancer surgery from January to December 2008. This study also included total 361 women (106 cases and 255 controls) who underwent screening test from February 2006 to December 2011 at Samsung Medical Center, Seoul, Korea. For this institution, cases were women who were diagnosed with invasive breast cancer in screening test. For each case of two institution, approximately two controls from same institution who had no evidence of lesions for at least one year after a screening test were randomly selected and matched for age at mammography (±1 years), menopausal status, and date of screening test (±1 month). This study was approved by institutional review board of Seoul National University (IRB No. 1902/003-014).

2.2 Measurement of mammographic density

Cranio-caudal view mammographic images were obtained using full-field digital mammography system (Senograph 2000D/Essential/DS, General Electric Company, Milwaukee, WI, USA). For cases, only the breast contralateral to that involved in the cancer diagnosis was used. For controls, breast side for measurement was randomly selected as in previous study [8]. All images were a vendor post-processed (i.e., "for presentation") image which was commonly used in clinical practice. Flowchart of data collection process is described in Figure 2.1.

The trained expert with more than 5 years of experience measured mammographic density using computer-assisted thresholding software, CUMULUS. The CUMULUS has widely been considered as gold-standard method of quantifying mammographic density [13, 14]. In this method, reader first choose a background threshold to segment breast region. Subsequently, reader choose a threshold (T_{dense}) to segment the dense region. Two quantitative measures of mammographic density were obtained; dense region (DA, cm^2) and percent density (PD, %). The DA was calculated by converting number of pixels belonging to dense region into cm^2. The PD was calculated as a ratio of the dense region in a breast region, multiplied by 100. I refer these two measures as *Cumulus*.

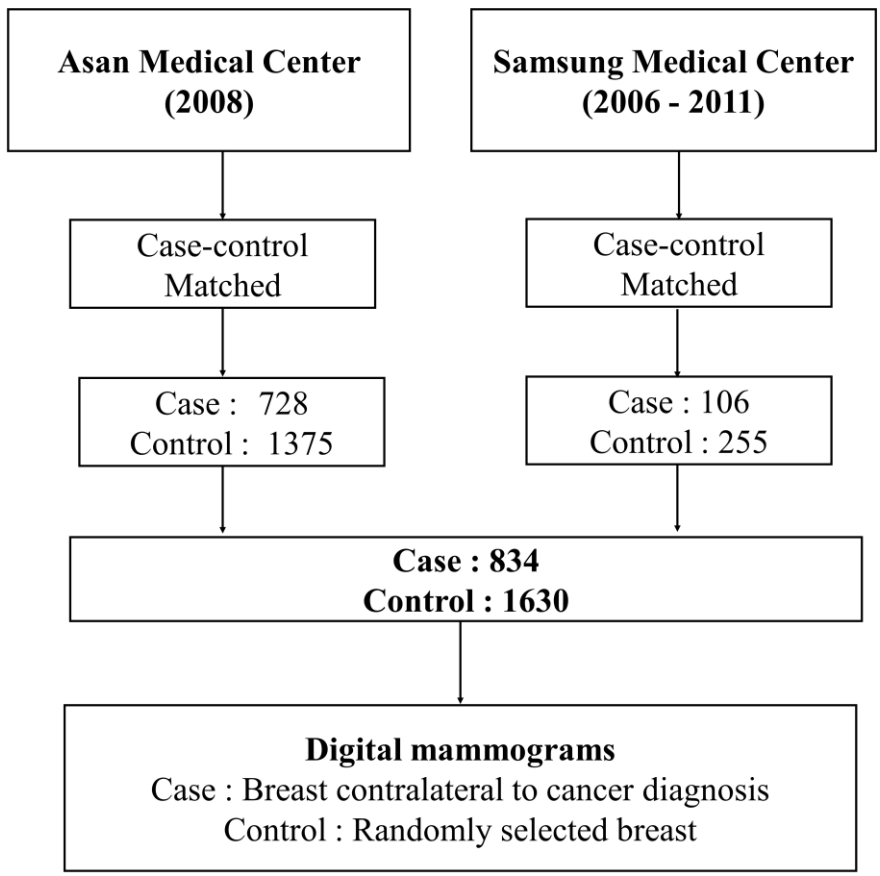

Figure 2.1. Flow chart of data collection process. First part of participants were from Asan Medical Center, Seoul, Korea (728 cases and 1375 controls, January to December 2008). Total 361 women (106 cases and 255 control, February 2006 to December 2011) were collected from Samsung Medical Center. For each case, approximately two controls from same institution were matched for age at mammography (±1 years), menopausal status, and date of screening test (±1 month).

2.3 Development of DL based algorithm

2.3.1 Establishing ground truth

All the participants were assigned to training set (n=1478), validation set (n=493) or test set (n=493). For training set, I augmented images with vertical and horizontal flips. Therefore, total 5912 images were prepared to train DL model. For images of training and validation set, I made corresponding binary images where each pixel value represents whether it belongs to dense tissue using T_{dense} (Method 2.2). These binary images were used as ground truths of mammographic density measurements.

2.3.2 Image preprocessing

All images were preprocessed with following two-steps. Firstly, the original images (2294 × 1914), were resized to available image size for DL model (224 × 256) using 2-D interpolation method. Secondly, the Contrast Limit Adaptive Histogram Equalization (CLAHE) was performed to enhance local contrast of images [15]. The CLAHE is a contrast enhancement technique which divides the image into multiple rectangular regions and then applies histogram equalization over each region. This method has been successfully applied to improve image contrast and increased the contrast between different types of tissues in digital mammograms [16-18]. Finally, total 5912 preprocessed images and corresponding binary images were prepared to establish the DL model.

2.3.3 Establishing DL model

DL model was established based on U-net [19]. U-net is a fully convolutional network and widely being used for biomedical image segmentation. For breast cancer research, U-net have been used for segmenting breast and fibroglanduar tissue in MRI volume [20], and detecting tissue lesion in digital mammogram [21]. U-net consists of a contracting path (left part of U) and an expansive path (right part of U). The contracting path can be seen

as down-sampling stage and follows the typical architecture of a convolutional network; convolution and pooling followed by Rectified Linear Units (ReLU). The expansive path can be seen as up-sampling stage. Every step in the expansive path consists of a up-convolution that halves the number of feature channels, a concatenation with the corresponding cropped feature map from the contracting path, and two convolutions, each followed by a ReLU.

Based on original U-net architecture, I added Drop-out and Batch Normalization layer after a convolutional layer to improve performance of the model. [22, 23]. The underlying architectures of DL model are shown in Figure 2.2 and Table 2.1. I chose hyperparameters for DL model from previous papers [10, 19], which are shown in Table 3.3. The Adaptive Moment Estimation (Adam) with learning rate of 0.1 was used as an optimization algorithm. I set batch size as 10 and applied early stopping to find optimal epoch based on performance on validation set (n=493). The mean intersect over union (IOU) was used to evaluate performance on validation set. DL model was implemented with Keras (2015, version 3.5.1; keras.io

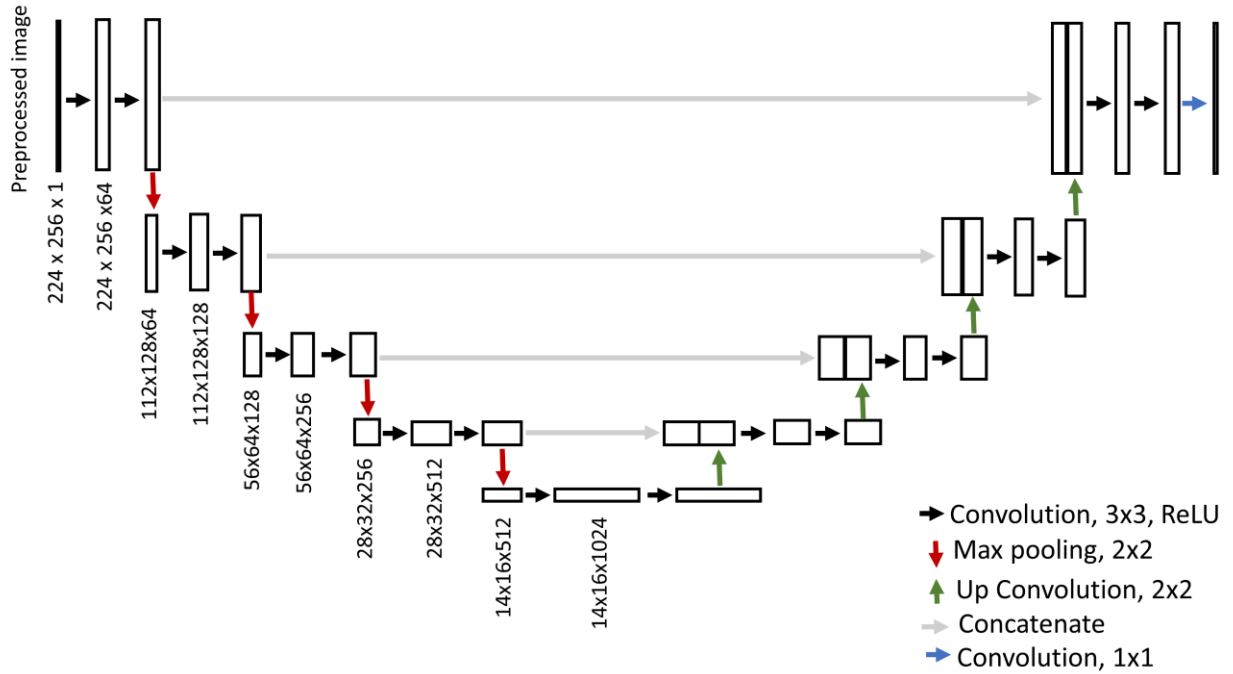

Figure 2.2. Architecture of DL model. Each box represents feature map where width indicates number of feature channel. The size of feature map is denoted under the each box (width, height and number of feature map). The colored arrows indicate different operations applied to each feature map. The output image is the same size as input image, where each pixel represents probability of belonging to a dense region.

Table 2.1. Architecture of DL model

Layer type	Input	Kernel size	Stride	#Repetition	Output size
Input image					256x224x1
Conv1[1]	Input image	3x3	1x1	2	256x224x64
BN1[2]	Conv1				256x224x64
Max pooling1	BN1	2x2	2x2	1	128x112x64
Conv2	Max pooling1	3x3	1x1	2	128x112x128
BN2	Conv2				128x112x128
Max pooling2	BN2	2x2	2x2	1	64x56x128
Conv3	Max pooling2	3x3		2	64x56x256
BN3	Conv3				64x56x256
Max pooling3	BN3	2x2	2x2	1	32x28x256
Conv4	Max pooling3	3x3	1x1	2	32x28x512
BN4	Conv4				32x28x512
Max pooling4	BN4	2x2	2x2	1	16x14x512
Conv5	Max pooling4	3x3	1x1	2	16x14x1024
BN5	Conv5				16x14x1024
Upconv1	BN5	2x2	1x1	1	32x28x512
Concatenate1	Upconv1 + BN 4			1	32x28x1024
Conv6	Concatenate1	3x3	1x1	2	32x28x512
BN6	Conv6				32x28x512
Upconv2	BN6	2x2	1x1	1	64x56x256
Concatenate2	Upconv2 + BN3			1	64x56x512
Conv7	Concatenate2	3x3	1x1	2	64x56x256
BN7	Conv7				64x56x256
Upconv3	BN7	2x2	1x1	1	128x112x128
Concatenate3	Upconv3 + BN2				128x112x256
Conv8	Concatenate3	3x3	1x1	2	128x112x128
BN8	Conv8				128x112x128
Upconv4	BN8	2x2	1x1	1	256x224x64

Concatenate4	Upconv4 + BN1				256x224x128
Conv9	Concatenate4	3x3	1x1	2	256x224x64
BN9	Conv9				256x224x64
Conv out	BN9	1x1	1x1	1	256x224x1

[1]Convolutional layers were followed by ReLU activation function.

[2]Batch normalization

2.3.4 Estimation of mammographic density

Entire processes of developing DL model and estimation process are described in Figure 2.3. Two quantitative measures of mammographic density, percent density (PD, %) and dense area (DA, cm^2) were estimated with following algorithm. For each mammographic image, at first, breast region was segmented using threshold which was determined by averaging two parameters of Gaussian Mixture Model (Figure 2.4). In parallel, each mammographic image was preprocessed as described in Method 2.3.2 and fed into DL model. The output of DL model was a score map between 0 and 1 which represent the probability that a given pixels belongs to dense tissue. On this score map, dense region was segmented by applying threshold at 0.5. Finally, DA was obtained by converting number of pixels in dense region to cm^2. PD was calculated as a fraction of pixels that was segmented as dense region in the breast region.

Table 2.2. Hyperparameters of DL model

Type	Hyperparameter	Value
CLAHE[1]	Clip limit	2
CLAHE	Grid size	4, 4
DL Model	Batch size	10
DL Model	Optimizer	Adaptive Moment Estimation (Adam)
DL Model	Optimizer – beta1[2]	0.9
DL Model	Optimizer – beta2	0.999
DL Model	Optimizer – learning rate	0.1
DL Model	Drop-out rate	0.5

[1]Contrast limit adaptive histogram equalization [15]

[2]Beta1 and beta2 are moment parameters for Adam optimizer [22]

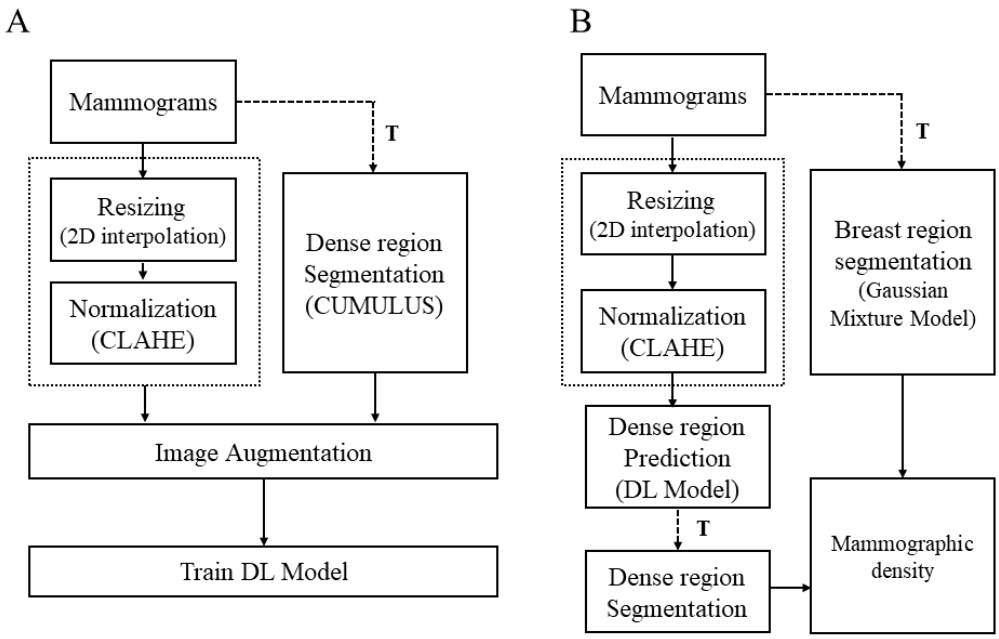

Figure 2.3 Diagrams of algorithm. Left diagram (A) describes training procedure of DL model. Right diagram (B) describes algorithm of estimating mammographic density using trained DL model. The dashed boxes indicate preprocessing step. The processes to segment regions were denoted as dashed arrows. T refers to the threholding to convert input image to masked image.

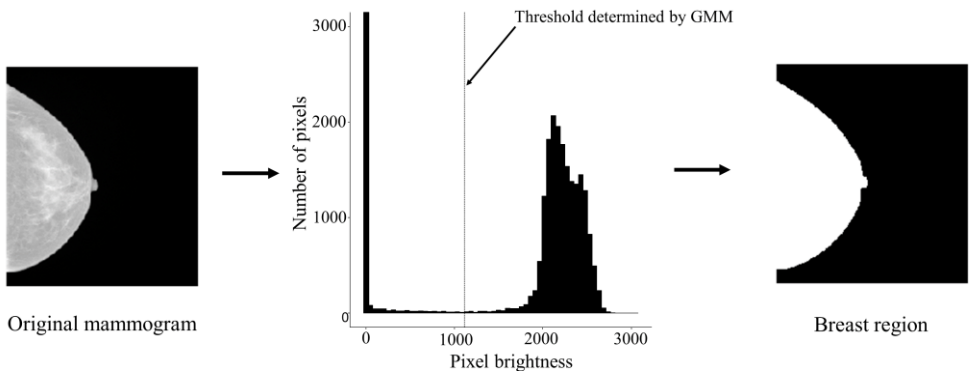

Figure 2.4. Breast region segmentation method. The histogram represents a distribution of the pixel brightness of an image. A vertical line indicates a threshold determined by averaging two parameters from Gaussian Mixture Model. All pixels that have larger brightness than this threshold are determined as a breast region.

2.4 Statistical methods

2.4.1 Agreement statistics

DL model was externally validated for agreement with the expert using held-out test of 493 women. Pearson's correlation (r), Scatter plot and Bland-Altman plot were used to assess agreement. To evaluate qualitative agreement of DL model, images were assigned to four groups based on the quartiles of each measurement and weighted kappa coefficient (κ) was estimated.

2.4.2 Evaluation of risk association

DL model should be validated for the purpose that is actually used. Thus, DL model was assessed as their ability to predict breast cancer. To make data approximately normal distribution, each of measures was transformed using Box-Cox power function. By using transformed measures of control participants, residuals of linear regression model after adjusting for age and BMI were calculated and then corresponding mean, μ and standard deviation, s of residuals were obtained. By using fitted model, residuals were standardized with μ and s. We refers to these measures as transformed, adjusted and standardized measures as described in previous papers [8]. From these measures, odds per adjusted standard deviation (OPERA) and area under receiver operating characteristic curve (AUC) were estimated using logistic regression. All the statistical analysis were performed using R, version 3.5.2 (R core development team, 2018).

Chapter 3

Results

3.1 Characteristics of study participants

The descriptive statistics of study participants by case-control status are shown in Table 3.1. *Cumulus* represent manual measures of experts that was obtained using CUMULUS software. Both PD and DA showed significant differences between case and control group. The aim of this study was to develop DL model which can automatically measure PD and DA and validate whether these differences are present.

Table 3.1. Characteristics of participants

	Control (n = 1630)	Case (n = 834)
Body mass index (kg/m^2), mean (SD)	22.6 (2.6)	23.2 (2.9)
Age at mammogram (yeas), mean (SD)	50.2 (5.7)	49.0 (5.5)
<40 (n, %)	5 (0.4%)	3 (0.4%)
40-49	618 (47.4%)	384 (57.6%)
50-59	646 (49.5%)	271 (40.6%)
> 60	35 (2.7%)	9 (1.3%)
Cumulus		
Percent density (%), mean (SD)	22.6 (13.8)	27.2 (15.1)
Dense area (cm^2) , mean (SD)	20.0 (14.0)	26.8 (17.2)

3.2 Agreement of DL model

The agreement statistics are shown in Table 3.2. The correlation coefficient was very high for both PD and DA (r = 0.94 and 0.96 for PD and DA, respectively). Scatter plots of DL measures and *Cumulus* are shown in Figure 3.2. An example of DL measures with corresponding manual measures are shown in Figure 3.1. In order to assess systematic differences between DL measures and *Cumulus*, the Bland-Altman plot analysis was carried out. The Bland–Altman plot is a scatter plot in which the difference between the two measures (A-B) is plotted against their mean value ([A+B]/2). For PD, as measures increased, DL model slightly underestimated mammographic density (mean difference, 2.95 %, Figure 3.3). In contrast, for DA, bias was negligible (mean difference, 0.4 %, Figure 3.3). There were several images for which DL model overestimated mammographic

density. An example of image showing large discrepancy between expert and DL model is shown in Figure 3.4.

Table 3.2. Agreement statistics

Measurements	Mean (SD)	Pearson's correlation (95% CI)	Mean absolute error, (Range)	Weighted κ coefficient[1] (95% CI)
Percent density (%)				
Cumulus	25.5 (14.9)	0.94	4.4	0.89
DL model	22.6 (12.6)	(0.93-0.95)	(0.05-30.84)	(0.87-0.91)
Dense area (cm^2)				
Cumulus	23.8 (16.4)	0.96	3.2	0.91
DL model	23.4 (15.7)	(0.95–0.97)	(0.01-35.93)	(0.89-0.93)

[1] Weighted kappa coefficients were computed using qualtile of each measure.

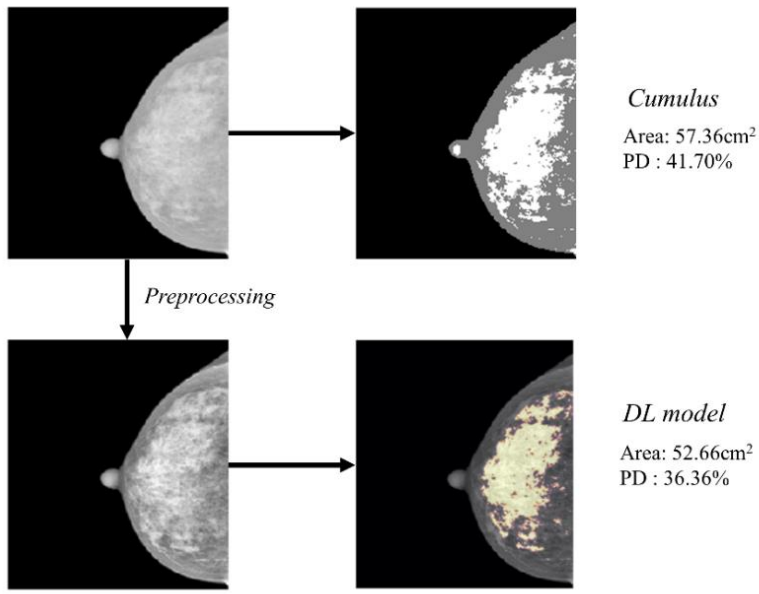

Figure 3.1. Mammographic density measurements. Top row images show manual measures using CUMULUS software. Top bottom images show DL measures from preprocessed image.

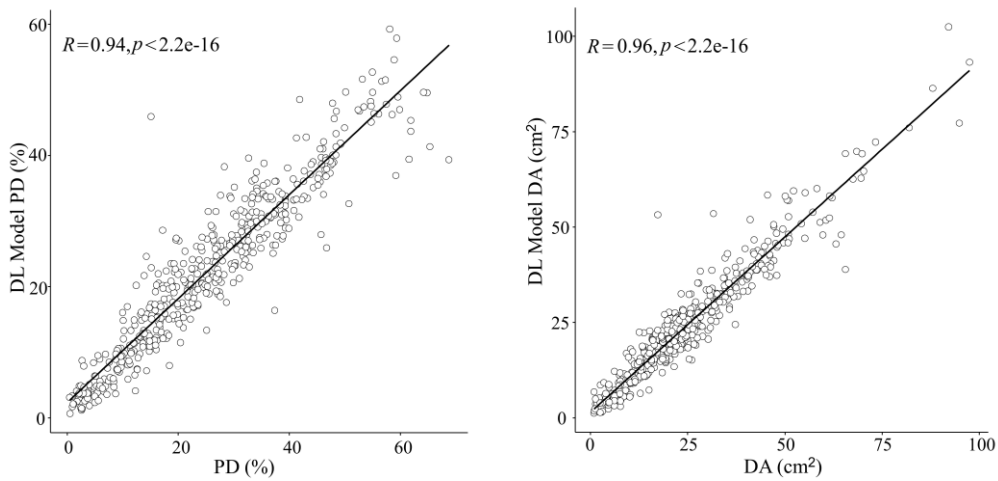

Figure 3.2. Scatter plots of DL model and expert. Left figure shows the scatter plot of PD between DL model and *Cumulus* (r = 0.94; 95% C.I 0.93 – 0.95,$y = 0.3 + 1.11x$). Right figures shows the scatter plot of DA between DL model and *Cumulus* (r = 0.96; 95% C.I 0.94–0.97).

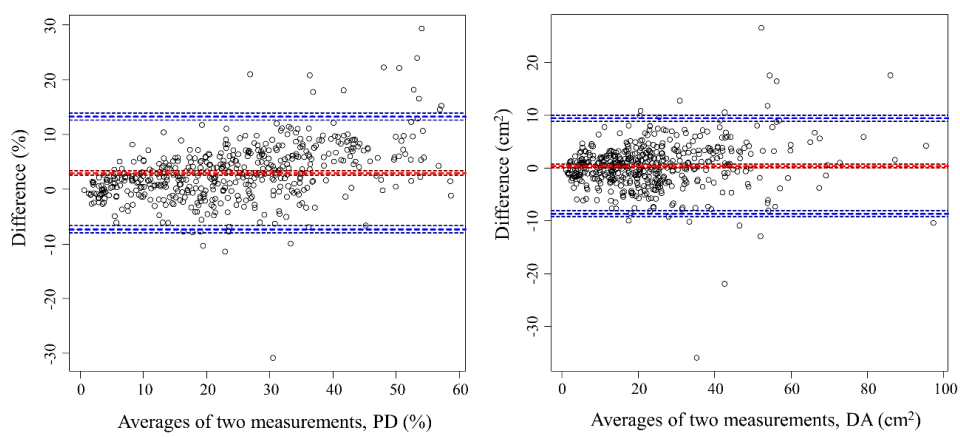

Figure 3.3. Bland-Altman plots of DL model and expert. Left and right figures show the Bland-Altman plot of PD and DA, respectively. Red dashed line indicates mean difference (2.95, 95% C.I 2.48 – 3.41 for PD and 0.4, 95% C.I 0 – 0.81 for DA). Blue dashed lines indicates lower and upper limits of agreement. (lower limit, -7.36, and upper limit, 13.24 for PD and lower limit, -8.63 and upper limit, 9.44 for DA, respectively).

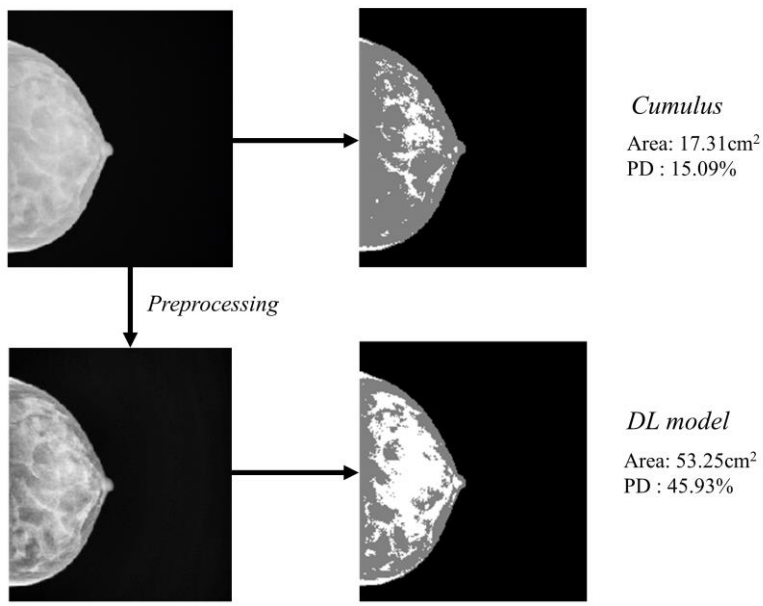

Figure 3.4. An example of outlier. This figure shows an example of image showing a large discrepancy between DL model and expert. Top row images show manual measures using CUMULUS software. Top bottom images show DL measures from preprocessed image.

3.3 Breast cancer risk profiles

The transformed standardized and adjusted measures were used to assess risk associations of measures adjusted for age and BMI. The risk profiles of all measures are shown in Table 3.3. When considering AUC and OPERA as measures of risk association, it was evident that DL measures were comparable to the manual measures of expert in terms of risk prediction. The differences of AUC were negligibly small (Absolute differences of AUC were 0.02 and 0.03 for PD and DA, respectively). However, risk associations across quartiles were not completely in agreement between *Cumulus* and DL model. For DL model, OR of very low density and high density decreased (Q3 vs. Q1, 0.46 and 0.54 for PD and DA, respectively) and the OR between very row density and very high density

slightly increased in both PD and DA (Q4 vs. Q1, 0.15 and 0.09 for PD and DA, respectively).

Table 3.3 Breast cancer risk profiles

Cumulus						
Percent density[1]	Cases (n)	Controls (n)	OR	95% C.I	P-value	AUC (95% C.I)
Q1[2]	26	82	1	-	-	
Q2	23	81	0.90	0.26 - 1.54	0.86	
Q3	53	81	2.06	1.50 - 2.63	< 0.05	**0.62**
Q4	65	82	2.50	1.95 - 3.05	< 0.01	**(0.56, 0.67)**
OPERA[3]	167	326	**1.51**	**1.25 - 1.83**	**< 0.001**	
Dense area[4]						
Q1	25	82	1		-	
Q2	25	81	1.01	0.38 - 1.65	1	
Q3	47	81	1.90	1.33 - 2.48	< 0.05	**0.63**
Q4	70	82	2.80	2.26 – 3.35	< 0.001	**(0.58, 0.68)**
OPERA	167	326	**1.61**	**1.34 - 1.95**	**< 0.001**	
DL Model						
Percent density						
Q1	26	82	1		-	
Q2	31	81	1.21	0.60 - 1.81	0.65	**0.61**
Q3	41	81	1.60	1.02 - 2.18	< 0.05	**(0.56, 0.67)**
Q4	69	82	2.65	2.11 - 3.20	< 0.001	
OPERA	167	326	**1.51**	**1.26 - 1.83**	**< 0.001**	
Dense area						
Q1	26	82	1		-	
Q2	31	81	1.21	0.60 - 1.81	0.65	**0.64**
Q3	35	81	1.36	0.77 – 1.96	0.38	**(0.58, 0.69)**
Q4	75	82	2.89	2.34 - 3.43	< 0.001	
OPERA	167	326	**1.63**	**1.36 - 1.98**	**< 0.001**	

[1] Percent density was adjusted for age and body mass index.

[2] Quartiles (Q1–Q4) were defined by distribution of measure adjusted for age and BMI.

[3] Odds ratio per standard deviation.

[4] Dense area was adjusted for age and body mass index.

Chapter 4

Discussion

I presented DL model for mammographic density assessment using extensive digital mammograms for which mammographic densities were quantitatively measured by expert. The DL model showed very good agreement with expert ($r = 0.94 – 0.96$ and $\kappa = 0.89 – 0.91$), and it was better than conventional method based on adaptive thresholding ($r = 0.63$) [26], clustering ($r = 0.82 – 0.85$) [27] and collective multiple measurements ($r = 0.89$) [6]. DL model was also competitive with other types of DL methods in previous papers, eg., 0.85 [9] and 0.81 [10].

The measures of DL model showed strong risk associations with breast cancer. It were comparable to results of previous studies conducted on Korean women (AUC, 0.6 [8] and 0.56 [28]) and Australian women (AUC, 0.62 – 0.64 [29]). It suggests that DL model is capable of reproducing mammographic density measures which are strong predictor of breast cancer.

One of the major advantage of DL model is that it is scalable method, because it learns features from data itself. Therefore, DL model can be easily adjusted to images of other modalities (e.g., ultrasound, magnetic resonance imaging or tomosynthesis) or density measures at higher thresholds, which were called Altocumulus and Cirrocumulus in previous papers [28, 29]. DL model may provide a way to measure different level of

mammographic densities in a reproducible and objective manner in the circumstances with images from various modalities and acquisition techniques are present.

The limitation of this study was that the measures from a single expert were considered “truth” of mammographic density. This could miss an opportunity to exceed performance of expert in terms of risk prediction. In future work, DL model that incorporates measures from multiple reader would be necessary. Since DL model could balance variability of measures between readers, it may contributes to minimizing subjectivity and will be closer to the “truth” as shown in a previous paper [12].

The possibility of developing personalized screening, taking mammographic density into account, is being widely discussed [30, 31]. However, inconsistency in assessment have been recognized to cause patient anxiety and could result in unnecessary screening test. DL model could potentially address concerns of current issue by reliably assessing mammographic density in a cost-effective way.

This study showed the potential of DL as a mammographic density assessment tool in clinical practice. In addition, DL model integrated with other risk factors was able to reliably assess breast cancer risk. If combined with other well known risk factors such as familiar causes and genetic marker, it could help clinicians inform patients of more precise breast cancer risk which takes mammographic density into account. Further research will be required to validate DL model across institutions and images from other modalities.

Bibliography

[1] J.N.Wolfe. Breast patterns as an index of risk for developing breast cancer. American Journal of Roentgenology. 1976. 126 (6):p. 1130–1139.

[2] Boyd N. F, et al. Mammographic densities and breast cancer risk. Cancer Epidemiology Biomarkers & Prevention. 1998. 7(12):p. 1133.

[3] McCormack V. A. and I. dos Santos Silva. Breast Density and Parenchymal Patterns as Markers of Breast Cancer Risk: A Meta-analysis. Cancer Epidemiology Biomarker and Prevention. 2006. 15(6):p. 1159.

[4] John L. Hopper. Odds per Adjusted Standard Deviation: Comparing Strengths of Associations for Risk Factors Measured on Different Scales and Across Diseases and Populations. American Journal of Epidemiology. 2015. 182(15):p. 863–867.

[5] Byng JW, et al. The quantitative analysis of mammographic densities. Phys Med Biol. 1994. 39 (10): p. 1629–1638.

[6] Li J, et al. High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer. Breast Cancer Research. 2012. 14(4):p. 114.

[7] Alonzo-Proulx O, et al. Reliability of Automated Breast Density Measurements. Radiology. 2015. 275(2):p. 366-376.

[8] Nguyen T. L, et al. Breast Cancer Risk Associations with Digital Mammographic Density by Pixel Brightness Threshold and Mammographic System. Radiology. 2017. 286(2):p. 433-442.

[9] M. Kallenberg et al. Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring. IEEE Transactions on Medical Imaging. 2016. 35(5):p. 1322-1331.

[10] Lee J and Nishikawa R. M. Automated mammographic breast density estimation using a fully convolutional network. Med. Phys. 2018. 45:p. 1178-1190.

[11] Mohamed A. A, et al. A deep learning method for classifying mammographic breast density categories. Med. Phys. 2018. 45:p. 314-321.

[12] D. Lehman, et al. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. Radiology. 2019. 290(1):p. 52-58

[13] S. Ourselin, et al. Automated volumetric breast density derived by shape and appearance modeling. Medical Imaging. 2014. 9034

[14] M.Nielsen, et al. Anovel and automatic mammographic texture resemblance marker is an independent risk factor for breast cancer. Cancer Epidemiology. 2011. 35(4):p. 381–387.

[15] Pizer SM. Psychovisual issues in the display of medical images. Pictoral Information Systems in Medicine. 1985. p 211-234

[16] N. Dhungel, et al. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med. Image Anal. 2017. 37(1):p. 114–128.

[17] E. Kozegar M, et al. Assessment of a novel mass detection algorithm in mammograms. J. Cancer Res. Ther. 2013:p. 592–600.

[18] N. Dhungel, G. Deep structured learning for mass segmentation from mammograms. Image Processing (ICIP). IEEE Inter- National Conference. 2015. p. 2950–2954.

[19] Ronneberger O, et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015. 9351:p. 234-241.

[20] Ufuk Dalmş, et al. Using Deep Learning to Segment Breast and Fibroglanduar Tissue

in MRI Volumes. Medical Physics. 2017. 44(2):p. 533-546.

[21] Moor Timothy de, et al, Automated Soft Tissue Lesion Detection and Segmentation in Digital Mammography Using a U-Net Deep Learning Network. IWBI , At Atlanta, Georgia. 2018.

[22] Ioffe, et al. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML. 2015.

[23] Srivastava Nitish, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research. 2014. 15:p. 1929−1958.

[26] C. Nickson Y.,et al, AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes. Breast Cancer Research. 2013. 15(5):p. 80.

[27] B. M. Keller, et al. Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. Medical Physics. 2012. 39(8):p. 4903–4917.

[28] Nguyen TL, et al. Mammographic density defined by higher than conventional brightness threshold better predicts breast cancer risk for full-field digital mammograms. Breast Cancer Res. 2015. 17(1)p. 142.

[29] Nguyen TL, et al. Mammographic Density Defined by Higher than Conventional Brightness Thresholds Better Predicts Breast Cancer Risk. 2017. 46(2):p. 652–661.

[30] Borgquist S, et al. Towards Prevention of Breast Cancer: What Are the Clinical Challenges? Cancer Prevention Research. 2018. 11(5):p. 255.

[31] Lee CPL, et al. Mammographic Breast Density and Common Genetic Variants in Breast Cancer Risk Prediction. PLOS ONE. 2015. 10(9):e0136650.

초록

유방 내 유방 실질 조직의 양을 반영하는 유방 밀도는 맘모그램에서 나타나는 밝은 부분으로 정의되며, 유방암의 강력한 위험인자로 널리 알려져 있다. 하지만 유방 밀도는 측정하는데 시간과 비용이 많이 든다는 단점으로 인해 유방암 검진 과정에서 제한적으로 사용돼 왔다. 본 연구의 목적은 유방암 검진에서 유방암 예측 모형에 포함해 활용할 수 있는 딥러닝 기반 유방 밀도 측정치를 개발하는 것이다.

본 연구는 아산 병원과 삼성 서울병원의 유방암 검진 자료로부터 수집된 총 2464 명의 참여자 (환자: 834 명, 대조군 : 1630 명) 를 대상으로 수행되었다. 환자의 경우 병변이 발생한 유방의 반대쪽 유방, 대조군의 경우 임의로 고른 유방을 대상으로 유방 밀도 측정에 5년 이상의 경력을 가진 전문가가 CUMULUS 프로그램을 활용하여 유방 밀도 (치밀 유방 부위, cm^2 및 치밀도 백분율, %) 를 측정하였다. 이 전문가 측정치를 훈련 데이터로 하여 완전 합성곱 신경망 (Fully Convolutional Network) 기반 딥러닝 모델을 구축하였고, 이를 테스트 데이터에 대해 적용해 전문가 측정치와의 일치도 및 유방암 예측력을 평가하였다.

딥러닝 모델은 전문가와 높은 일치도 ($r = 0.94 - 0.96$, weighted $\kappa = 0.89 - 0.91$) 를 보였다. 또한 나이와 BMI를 보정한 딥러닝 기반 측정치의 유방암 예측력을 평가한 결과, 딥러닝 모델이 전문가와 비슷한 수준의 예측력을 갖는다는 것을 확인하였다 (전문가, AUC = 0.62 – 0.63, 딥러닝 모델, AUC = 0.61 – 0.64).

본 연구는 딥러닝이 현재의 노동 집약적인 유방 밀도 측정법을 보완할 수 있는 가능성을 보여주었다. 이는 비용-효율적인 방법으로 유방 밀도

측정치를 유방암 예측 모형에 포함시킬 수 있는 기회를 제공한다. 이러한 맘모그램 기반 유방암 위험도 예측 모형이 유방암 검진 과정에 적용된다면 보다 정밀한 유방암 위험도 평가를 통해 효과적으로 유방암 고위험군을 선별할 수 있으며, 고위험군에 대한 맞춤형 예방 전략이 적용된다면 장기적으로 유방암 조기 발견 및 사망률 감소에 기여할 수 있을 것으로 기대한다.

주요어 : 유방암, 유방 밀도, 유방암 예측 모형, 디지털 맘모그램, 딥러닝

학번 : 2017-27525